# Notes 18 Sampling spanning trees by random walk

## 1. Random spanning trees

Fix a connected graph $G = (V, E)$ on $n$ vertices. A spanning tree in $G$ is an acyclic subgraph of $G$ containing $n - 1$ edges.

We want to sample a spanning tree of $G$, (nearly) uniformly at random, as follows.

Random walk on spanning trees

Let $T_0$ be an arbitrary spanning tree of $G$
For $t = 0, 1, 2, \ldots$
    Remove an edge from $T_t$ uniformly at random to obtain $F_t$
    Among all spanning trees containing $F_t$, uniformly pick one as the new $T_{t+1}$

This is a random walk/Markov chain on an auxiliary weighted graph $\mathcal{T}_G$, whose nodes are spanning trees in $G$, and two nodes in $\mathcal{T}_G$ are adjacent if they share exactly $n - 2$ edges.

For decades, this random walk was conjectured to mix in polynomial time. It was recently proved by Anari, Liu, Oveis Gharan, and Vinzart.

**Theorem 1.1.** *The above random walk has eigenvalue gap at least* $1/(n-1)$.

Eigenvalue gap $\beta$ is the difference $\lambda_1 - \lambda_2$ between the two largest eigenvalues. By results in Notes12, the lazy version of the random walk mixes in polynomial time. Recall that the lazy random walk mixes in time $O((\log |V(\mathcal{T}_G)|)/\beta)$. Since $|V(\mathcal{T}_G)| \leqslant \binom{\binom{n}{2}}{n-1} \leqslant \binom{n^2}{n-1} \leqslant n^{2(n-1)} = \exp(O(n \log n))$, the lazy random walk mixes in time $O(n^2 \log n)$. (In fact the usual, non-lazy random walk also has the same mixing time bound.)

## 2. Bipartite inclusion graphs

Since the ambient graph $G$ is fixed, we identify a spanning tree $T$ in $G$ with the set of $n-1$ edges in $T$.

Denote by $Y(n-1)$ the sets of edges of spanning trees in $G$. More generally, define

$$Y(k) = \left\{ F \in \binom{E}{k} \mid F \text{ is acyclic} \right\} \qquad \text{for } 0 \leqslant k \leqslant n-1 \,.$$

Decompose the random walk over spanning trees into two transitions:

- (Down) Go from $T_t \in Y(n-1)$ to $F_t \in Y(n-2)$ by removing an edge uniformly at random
- (Up) Then from $F_t$ to $T_{t+1} \in Y(n-1)$ by choosing $T_{t+1}$ uniformly from among all $T_{t+1} \supset F_t$

Down transition corresponds to matrix $D_{n-1}$, which is $Y(n-1)$-by-$Y(n-2)$ (recall that a row probability vector multiplies on the left to a transition matrix). Likewise up transition corresponds to matrix $U_{n-1}$, which is $Y(n-2)$-by-$Y(n-1)$.

$D_{n-1}U_{n-1}$ is the transition matrix for the random walk. We want to bound $\beta(D_{n-1}U_{n-1})$.

We think of up and down transitions as random walk transitions on the auxiliary graph $\Gamma_{n-1}$:

**Definition 2.1** (Bipartite inclusion graph)**.** For $0 \leqslant k \leqslant n-1$, $\Gamma_k$ has vertex set $Y(k) \cup Y(k-1)$. $F \in Y(k)$ is adjacent to $F' \in Y(k-1)$ if $F \supset F'$.

We will look at $\Gamma_k$ for $0 \leqslant k \leqslant n-1$ later on to apply induction.

In $\Gamma_k$, down transition means moving from $F \in Y(k)$ to a random neighbor $F' \in Y(k-1)$, uniformly from among all $k$ neighbors of $F$.

We will define up transition matrix $U_k$ shortly, to move from $F' \in Y(k-1)$ to a random neighbor $F \in Y(k)$ from some distribution.

Thanks to the following proposition, $D_{n-1}U_{n-1}$ and $U_{n-1}D_{n-1}$ have the same eigenvalue gap, whenever this gap is at most 1.

**Proposition 2.2.** *Given any matrices $A$ and $B$, $AB$ and $BA$ have the same non-zero eigenvalues with the same multiplicities.*

This proposition can be proved by showing $AB$ and $BA$ have essentially the same characteristic polynomials. Look up "Characteristic polynomials" on Wikipedia if interested.

Kaufman and Oppenheim came up with a way to relate $\beta(U_{n-1}D_{n-1})$ to $\beta(D_{n-2}U_{n-2})$. Of course, $\beta(D_{n-2}U_{n-2}) = \beta(U_{n-2}D_{n-2})$. One can then inductively bound $\beta(U_kD_k) = \beta(D_kU_k)$.

**Proposition 2.3** (Kaufman–Oppenheim)**.** *$\beta(D_kU_k) \geqslant 1/k$ for $1 \leqslant k \leqslant n-1$.*

This main proposition implies the main theorem.

## 3. Matrices of conditional probabilities and their eigenvalue gaps

To bound $\beta(D_kU_k)$ in Proposition 2.3 for each fixed $k$, Kaufman and Oppenheim do another induction (as we will see later) and study, for every $F \in Y(0) \cup \cdots \cup Y(n-1)$, the matrix of conditional probabilities

$$A_F \stackrel{\text{def}}{=} \Big(\mathbb{P}[e \in T \text{ and } f \in T \mid F \subseteq T]\mathbb{1}(e \neq f)\Big)_{e,f \in E\setminus F}.$$

To get some intuition about these matrices of conditional probabilities, take $F = \emptyset$. Then

$$A_\emptyset = \Big(\mathbb{P}[e \in T \text{ and } f \in T]\mathbb{1}(e \neq f)\Big)_{e,f \in E}$$

encodes the probability that every pair of edges both appear in a random spanning tree.

The adjacency matrix $A_F$ has a normalized version $\mathcal{A}_F$, whose eigenvalue gap $\beta(\mathcal{A}_F)$ measures how random the edges are in a random spanning tree $T$, after conditioning on $F$ appearing in $T$. The closer this gap is to 1, the closer for every edge in $T$ to be uniform (from among those that can be added to $F$ to form a spanning tree).

To help compute the above conditional probabilities, we will look at

$$\sigma(F) \stackrel{\text{def}}{=} \mathbb{P}[\text{random spanning tree } T \text{ in } G \text{ contains } F].$$

For example, given distinct edges $e$ and $f$ not in $F$, the entry $A_F(e,f)$ equals $\dfrac{\sigma(F \cup \{e,f\})}{\sigma(F)}$.

## 4. $n$-bipartite inclusion graph

We can visualize using the $n$-bipartite inclusion graph $\Gamma_{n-1} \cup \cdots \cup \Gamma_1$. Its vertex set is $\mathcal{F} = Y(0) \cup \cdots \cup Y(n-1)$. In this and subsequent sections, we denote vertices in this graph with lowercase letters such as $t$ and $b$, as opposed to uppercase letters such as $F$ in the previous section.

Let's focus on a particular layer $Y(k)$. Sampling a random forest $f$ from $Y(k)$ with probability proportional to its probability mass $\sigma(f)$ gives us a distribution $\pi_k$ on $Y(k)$, so that $\pi_k(f) = \sigma(f)/\sum_{g \in Y(k)} \sigma(g)$. The distribution $\pi_k$ coincides with the following:

(1) Pick a random spanning tree $T$ from $G$, then randomly remove all but $k$ edges from $T$.

The probability mass function $\pi_k$ satisfies

$$\pi_k(f) = \frac{1}{k+1} \sum_{g \in Y(k+1), g \supset f} \pi_{k+1}(g) \tag{2}$$

because the random process in Eq. (1) is the same as the following:

(i) Choose a random spanning tree $T$ from $G$
(ii) Randomly remove all but $k+1$ edges from $T$ to obtain $g$
(iii) Randomly remove one more edge from $g$ to obtain $f$

Every $g \in Y(k+1)$ containing $f$ has probability $\dfrac{1}{k+1}$ to give rise to $f$.

For some applications, one may want to sample from a nonuniform target distribution $\pi$ over spanning trees in $Y(n-1)$. For example, $G$ may have edge weights $w_G$, and a natural distribution $\pi$ on spanning trees $T$ would be proportional to the product of edge weights in $T$, so that $\pi(T) \propto \prod_{e \in T} w_G(e)$.

Almost everything in our analysis still holds with an arbitrary distribution $\pi$ of spanning trees over $Y(n-1)$.

## 5. Up-walk and down-walk

In $\Gamma_k$, layer $k$ of the $n$-partite inclusion graph, there are two natural two-step walks:

- Up-walk $P^{\wedge}_{k-1} = U_k D_k$ on a weighted graph on vertex set $Y(k-1)$
- Down-walk $P^{\vee}_k = D_k U_k$ on a weighted graph on vertex set $Y(k)$

Focusing on $\Gamma_k$, abbreviate

$$T = Y(k) \qquad B = Y(k-1)$$

The down transition $D_k$ is given by

$$D_k(t,b) \stackrel{\text{def}}{=} \frac{1}{k}\mathbb{1}(b \subset t) \qquad \text{for } t \in T, b \in B\ .$$

Using Eq. (2), the up transition is given by

$$U_k(b,t) \stackrel{\text{def}}{=} \frac{\pi_k(t)}{k\pi_{k-1}(b)}\mathbb{1}(b \subset t) \qquad \text{for } b \in B, t \in T\ .$$

Then $\pi_{k-1}^{\top} U_k = \pi_k^{\top}$. This equality means the random process in Eq. (1) is equivalent to the following:

(i) Pick a random spanning tree $T$ from $G$
(ii) Randomly remove all but $k-1$ edges from $T$ to get $b$
(iii) Randomly add an edge outside $b$ to get a forest $t$ with probability proportional to $\pi_k(t)$

5.1. **Down-walk.** $P^{\vee}_k = D_k U_k$ induces a random walk on a weighted graph on $T$.
A step in $P^{\vee}_k$ also corresponds to a length-2 walk in $\Gamma_k$, from $T$ to $B$ to $T$.

5.2. **Up-walk.** $P^{\wedge}_{k-1} = U_{k-1} D_{k-1}$ corresponds to the random walk on a weighted graph on $B$.
A step in $P^{\wedge}_{k-1}$ also corresponds to a length-2 walk in $\Gamma_k$, from $B$ to $T$ to $B$.
An up-walk from any vertex $b \in B$ has probability $1/k$ of staying there, so that $P^{\wedge}_{k-1}(b,b) = 1/k$, because no matter which intermediate vertex in $T$ is visited by this length-2 walk, there is exactly probability $1/k$ of finally returning to $b$.

5.3. **Non-lazy up-walk.** $\tilde{P}^{\wedge}_{k-1}$ is the non-lazy version of $P^{\wedge}_{k-1}$.
A step in $\tilde{P}^{\wedge}_{k-1}$ also corresponds to a length-2 *path* in $\Gamma_k$, from $B$ to $T$ to a different vertex in $B$.

$$P^{\wedge}_{k-1} = \frac{1}{k} I + \frac{k-1}{k}\tilde{P}^{\wedge}_{k-1}\ .$$

$\pi_{k-1}$ is stationary for both $P^{\wedge}_{k-1}$ and $\tilde{P}^{\wedge}_{k-1}$, as a common left-eigenvector with eigenvalue 1.

## 6. Spectral comparison

We now upperbound the spectrum of $\tilde{P}^{\wedge}_k$ by the spectrum of $P^{\vee}_k$. It is more convenient to first transform each of them into a symmetric matrix with the same spectrum.

$\tilde{P}^{\wedge}_k$ and $P^{\vee}_k$ share a common stationary distribution $\pi_k$. These transitions are both of the form $P(b,b') = A_P(b,b')/\pi_k(b)$ for symmetric $A_P$, so $P = \Pi^{-1} A_P$, where $\Pi = \mathrm{Diag}(\pi_k)$. They both represent a random walk with adjacency matrix $A_P = \Pi P$ and common degree matrix $\Pi$.

**Definition 6.1.** Given matrices $P, Q, \Pi$, if $\Pi P, \Pi Q$ are symmetric, we write

$$P \preccurlyeq_{\Pi} Q \qquad \Longleftrightarrow \qquad \Pi P \preccurlyeq \Pi Q\ .$$

$P \preccurlyeq_{\Pi} Q$ is equivalent to $\Pi^{1/2} P \Pi^{-1/2} \preccurlyeq \Pi^{1/2} Q \Pi^{-1/2}$, if $\Pi$ is symmetric and $\Pi \succ 0$.
Since $P$ is similar to the symmetric matrix $\mathcal{A}_P = \Pi^{1/2} P \Pi^{-1/2}$ (the normalized adjacency matrix), they have the same spectra.
In our application, $P$ and $Q$ often represent transitions with a common stationary distribution (the main diagonal of $\Pi$). $\mathcal{A}_P$ and $\mathcal{A}_Q$ will have the same top eigenspace (spanned by $\sqrt{\pi}$, with eigenvalue 1). Applying Courant–Fischer to the orthogonal subspace,

$$P \preccurlyeq_{\Pi} Q \qquad \Longleftrightarrow \qquad \mathcal{A}_P \preccurlyeq \mathcal{A}_Q \qquad \Longrightarrow$$

$$\lambda_2(\mathcal{A}_P) = \sup_{\substack{x \perp \sqrt{\pi} \\ \|x\|=1}} x^{\top} \mathcal{A}_P x \leqslant \sup_{\substack{x \perp \sqrt{\pi} \\ \|x\|=1}} x^{\top} \mathcal{A}_Q x = \lambda_2(\mathcal{A}_Q)\ ,$$

and hence $\lambda_2(P) = \lambda_2(\mathcal{A}_P) \leqslant \lambda_2(\mathcal{A}_Q) = \lambda_2(Q)$.

**Proposition 6.2.** *For $1 \leqslant k \leqslant n-2$, let $\Pi_k = \mathrm{Diag}(\pi_k)$. Then $\tilde{P}_k^{\wedge} \preccurlyeq_{\Pi_k} P_k^{\vee}$.*

This proposition is proved in the next section. We now show how it implies Proposition 2.3.

*Proof of Proposition 2.3.* $\lambda_1(P_k^{\vee}) = 1$. We prove by induction that $\lambda_2(P_k^{\vee}) \leqslant 1 - \frac{1}{k} = \frac{k-1}{k}$.

When $k = 1$, $P_1^{\vee}(t,t') = \pi_1(t')$, so $P_k^{\vee} = \mathbb{1}\pi_1^{\top}$ has rank 1, and $\lambda_2(P_k^{\vee}) = 0 \leqslant \frac{k-1}{k}$.

For $k > 1$, $\tilde{P}_{k-1}^{\wedge} \preccurlyeq_{\Pi_k} P_{k-1}^{\vee}$ by Proposition 6.2, so $\lambda_2(\tilde{P}_{k-1}^{\wedge}) \leqslant \lambda_2(P_{k-1}^{\vee}) \leqslant \frac{k-2}{k-1}$.

Also $\lambda_2(P_{k-1}^{\wedge}) = \frac{1}{d} + \frac{k-1}{k}\lambda_2(\tilde{P}_{k-1}^{\wedge}) \leqslant \frac{k-1}{k}$.

$P_{k-1}^{\wedge} = U_k D_k$ and $P_k^{\vee} = D_k U_k$ share the same non-zero eigenvalues, thus $\lambda_2(P_k^{\vee}) \leqslant \frac{k-1}{k}$. □

## 7. Garland's method

We now discuss Proposition 6.2 that bounds $\tilde{P}_k^{\wedge}$ by $P_k^{\vee}$.

Focusing on adjacent layers $\Gamma_{k+1} \cup \Gamma_k$, abbreviate

$$T = Y(k+1) \qquad M = Y(k) \qquad B = Y(k-1) \qquad (\text{“top” “middle” “bottom”})$$

Garland method decomposes transitions of $P_k^{\vee}$ and $\tilde{P}_k^{\wedge}$ into unions of subgraphs.

7.1. **Down-walk.** A step in $P_k^{\vee} = D_k U_k$ represents a length-2 walk $(m, b, m')$ from $M$ to $B$ to $M$. Decompose the set of all such walks (over all possible starting vertex $m \in M$) based on the bottom vertex $b$.

Walks with the same $b$ corresponds to transitions on a weighted clique (with self-loops) over $S_b \subseteq M$, where

$$S_b = \{m \in M \mid m \supset b\} .$$

7.2. **Non-lazy up-walk.** A step in $\tilde{P}_k^{\wedge}$ represents a length-2 path $(m, t, m')$ from $M$ to $T$ to a different vertex in $M$. Decompose the set of all such walks (over all possible starting vertex $m \in M$) based on the common intersection $b = m \cap m'$ of this path.

Walks with the same $b$ corresponds to transitions on the weighted graph $H_b = (S_b, E_b)$ over $S_b$ with edge set

$$E_b = \{(m, m') \in S_b \times S_b \mid m \cup m' \in T\} .$$

7.3. **Spectra.** The down-walk and non-lazy up-walk are now decomposed into subgraphs on $S_b$, over various $b \in B$.

Let $\pi_b$ denotes the distribution of picking $t \sim S_b$ conditioned on $t \supset b$, and $\Pi_b = \mathrm{Diag}(\pi_b)$.

For $P_b^{\vee}$, the subgraph (with adjacency matrix $\Pi_b P_b^{\vee}$) is a clique with self-loops.

For $\tilde{P}_b^{\wedge}$, the subgraph $H_b$ (with adjacency matrix $\Pi_b \tilde{P}_b^{\wedge}$) has its edges determined by $Y(k+1)$.

Kaufman and Oppenheim proved that Proposition 6.2 still holds “when restricted to these two subgraphs on $S_b$”.

**Lemma 7.1.** *For $1 \leqslant k \leqslant n-2$ and $b \in Y(k-1)$, let $\Pi_b = \mathrm{Diag}(\pi_b)$. Then $\tilde{P}_b^{\wedge} \preccurlyeq_{\Pi_b} P_b^{\vee}$.*

This lemma is proved in the next lecture. Summing over $b \in Y(k-1)$ yields Proposition 6.2.

For $P_b^{\vee}$: It has rank 1 since $\Pi_b P_b^{\vee} = \mu_b \mu_b^{\top} / \pi_{k-1}(b)$ ($\mu_b$ is the vector of edge weights incident to $b$).

Therefore $\lambda_2(P_b^{\vee}) = \cdots = \lambda_n(P_b^{\vee}) = 0$, and $\Pi_b P_b^{\vee}$ is a clique that mixes perfectly in one step.

For $\tilde{P}_b^{\wedge}$: Turns out $\tilde{P}_b^{\wedge}$ and $P_b^{\vee}$ share the same top eigenvector with the same eigenvalue.

Therefore $\tilde{P}_b^{\wedge} \preccurlyeq_{\Pi_b} P_b^{\vee}$ is equivalent to $\lambda_2(\tilde{P}_b^{\wedge}) \leqslant 0$.

In particular $\tilde{P}_b^{\wedge}$ must be a weighted expander.

## 8. Variations

Suppose you want apply the same random walk algorithm to a different setting, such as

- Uniformly sample a path of length $d$ in a graph $G$; or
- Uniformly sample a clique of size $d$ in a graph $G$.

Does the same random walk mix quickly in these settings?

All the constructions still make sense in those settings ($(d+1)$-partite inclusion graph, up- and down-walks, Garland's decomposition). But Lemma 7.1 may or may not hold.

Turns out $P_b^\vee$ from the previous section will still has rank 1 and represents a perfectly mixing weighted clique with $\lambda_2(P_b^\vee) = \cdots = \lambda_n(P_b^\vee) = 0$.

But now $\tilde{P}_b^\wedge$ may not represent weighted expanders and may violate $\lambda_2(\tilde{P}_b^\wedge) \leqslant 0$, because the subgraphs $H_b$ depend crucially on $Y(k+1)$ (and also indirectly on distribution $\pi$ at the top level).

## 9. Matroid

One general situation in which Lemma 7.1 holds (and hence fast mixing) is when the set system $Y(0) \cup \cdots \cup Y(d)$ is a matroid.

A matroid is a family $\mathcal{I}$ of subsets over a ground set $U$ that is:

(1) Nonempty: $\mathcal{I} \neq \emptyset$
(2) Downward closed: If $A \in \mathcal{I}$ and $B \subseteq A$, then $B \in \mathcal{I}$
(3) Exchangable: If $A, B \in \mathcal{I}$ and $|A| > |B|$, then there is $e \in A \setminus B$ such that $B \cup \{e\} \in \mathcal{I}$

Turns out all maximal $A \in \mathcal{I}$ have the same size $d$ (called the rank of the matroid).

The family of acyclic edges in a graph $G$ is an example of a matroid. The ground set $U$ is the set $E$ of edges in $G$. A subset $F \subseteq U$ belongs to $\mathcal{I}$ if $F$ is acyclic. Maximal $F \in \mathcal{I}$ are spanning trees in $G$. This matroid has rank $n-1$, where $n$ is the number of vertices in $G$. Fast mixing over spanning trees is thus a special case.